BOYD, PATRICIA SHAQIRAH, M.S. Computational Modeling of Small Non-Coding RNA Intramolecular Structures and the Determination of Their Collisional Cross Sections. (2018)
Directed by Norman Chiu. 38pp.

Ribonucleic acid (RNA) plays important and fundamental roles in different biological activities which include transmission of genetic information, regulation of gene expression, catalysis for biochemical reactions and biomarkers for different diseases[1]. The roles they play, as well as the interactions they perform within the body, are made possible by the stable secondary RNA structures that exist in the cell. The problem however, is that even through years of experimental work to determine the structures of RNA, little is known, and the majority of RNA structures remain structurally uncharacterized. RNA may undergo different types of modifications in the cell that can potentially impact its secondary structure. In addition to having limited knowledge on the actual RNA structures, none of current analytical methods can be universally used to detect all types of RNA modifications and its corresponding position that exist within a biological system. To address these issues for RNA research work as well as improving the accuracy on identifying specific RNA biomarkers, this study **aims** to investigate some of the intrinsic properties of RNA biomarkers that may affect the ion mobility mass spectrometric measurements of RNA samples. The **long-term goal** of this study is to develop an analytical method for differentiating and identifying isomeric RNA with or without any modification. Particularly, we are interested in the identification of specific isomeric RNA biomarkers which have

identical nucleotide composition and high sequence similarity. MicroRNA (miRNA) are selected as an initial model in this study

COMPUTATIONAL MODELING OF SMALL NON-CODING RNA INTRAMOLECULAR STRUCUTRES AND THE DETERMINATION OF THEIR COLLISIONAL CROSS SECTIONS

by

Patricia Shaqirah Boyd

A Thesis Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Greensboro
2018

Approved by

_____
Committee Chair

APPROVAL PAGE

This thesis written by PATRICIA SHAQIRAH BOYD has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair _____

Committee Members _____

_____

_____

_____
Date of Acceptance by Committee

_____
Date of Final Oral Examination

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER I

INTRODUCTION

1.1 Review of Literature

1.1.1 RNA

In 1868 Friedrich Miescher discovered nucleic acids, the building blocks of the biological molecules. A few decades later, Francis Crick predicted the functional ribonucleic acid components that mediated translation[2,3]. Ribonucleic acid (RNA) is a nucleic acid that is essential for different biological jobs like coding, decoding, regulation, gene expression. RNA are built with the four base ribonucleotides: guanosine (G), uridine (U), adenosine (A), and cytidine (C). These four bases are considered to be the four canonical ribonucleotides.

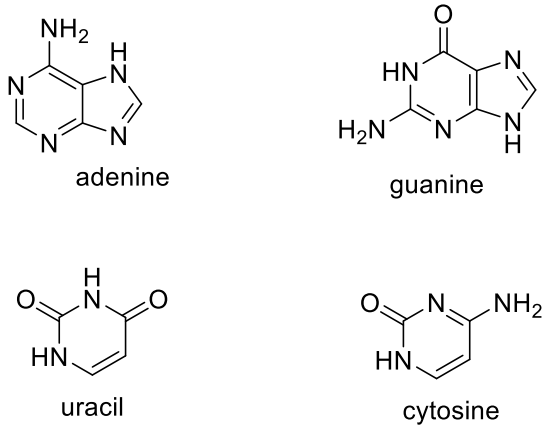

Figure 1. The Four Canonical Bases for RNA

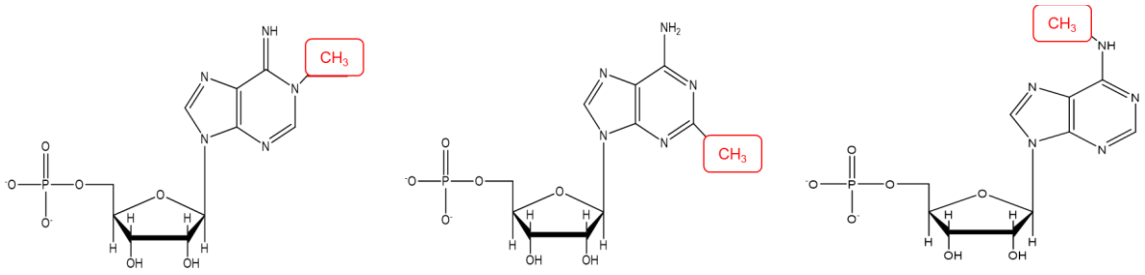

Figure 2. Modified Adenosine Base Pairs.

The four canonical bases are an important distinction because like in figure 1, these bases can undergo many different modifications that can impact the folding and the function of the RNA such as the modifications that are noted in figure 2

The other major macromolecule that comes from nucleic acids is deoxyribonucleic acids (DNA). Structurally, RNA and DNA are very similar but there are some key differences which cause their respective functions to differ greatly. One of the most major differences are their base pairs. DNA has the base thymine, while RNA has uracil. Another major difference that is displayed is the major and minor groves. DNA has smaller major and minor grooves than RNA. These differences are so large that is there two structures were to be aligned, they would not be superimposable. Because their function is dependent on their folding, the structurally difference between the two is key. DNA has the function of being in charge for the genome. RNA however is mainly in charge of decoding the information that is given on the genome. The DNA sequences rely on the double helix created by hydrogen bonding between two base pairing. Although RNA lays in a single strand, it will create hair pin loops and folds due to

the Watson-Crick base paring and the hydrogen bonding. This is pertinent as there are only four base pairs however, there are over 100 different types of modifications that they can undergo which allows an increase in the diversity of RNA structure and function[4].

1.1.2 Types of RNA

There are many different types of RNA, the most biologically active being mRNA, tRNA, rRNA, snRNA, and non-coding RNAs. Messenger RNA (mRNA) are used to convey the genetic information to aid in the production of proteins. They work in direct correlation with the genome in making different proteins for the body. Even viruses have their genetic information encoded by these RNAs. Transfer RNA (tRNA) are used to deliver amino acids to the ribosome were the ribosomal RNA (rRNA) link the amino acids together to form these proteins[5,6]. At the moment, these are the most heavily studied types of RNA because of their abundance and size. The primary focus of this project however, is non-coding RNA.

1.1.3 Noncoding RNA

Specifically, this project focuses on non-coding RNA. Non-coding RNA (ncRNA) are RNA molecules that do not translate into a protein. Despite this, there are many types of ncRNA which are responsible for regulating gene expression and associated with diseases such as cancer, cardiovascular, neurological, and metabolic diseases. There have been numerous reports of down regulation of ncRNA in tumors in comparison to normal tissues[7].

Depending on the size, these ncRNA are divided into two major groups- small and long ncRNA. Among the small ncRNA, miRNA, short interfering RNA and piwi-interacting RNA are studied in more detail. MiRNA has an average size of 22 nucleotides, and can be found in plants, animals and some viruses

1.1.4 MicroRNA

MicroRNA (MiRNA) are in the class of short non-coding RNA that post transcriptionally regulate gene expression. MiRNA are believed to regulate as much as 60% of all gene expression in humans. Although these are a newer class of RNA, they have been heavily reported as biomarkers for various diseases[4]. Additionally, they have also been used as potential drug targets. When looking at all of the known miRNA targets from miRBase, miRbase was accessed on 19 August 2016. it was seen that 55% of these RNA were structurally isomeric (meaning that they had the same size and identical nucleotide composition), with quite a few having high homology.

Currently, there are a total of 2588 human miRNA that are available in the miRbase. These miRNA range in size from 16 to 28 nucleotides, thus the average size of the human miRNA is 22 nucleotides. These miRNA can potentially be used for early detection of diseases and the creation of better diagnostics in the future.

1.1.5 Isomeric RNA

Different RNA molecules that have the same size and identical nucleotide composition are considered structural isomers. The structural isomers of miRNA are abbreviated “SimiR”. The isomeric human miRNA are categorized by the number of isomers present. Figure 3 shows that there are pairs of isomers which have as few as two structural isomers and those with as many as 13 different isomers. This depicts the distribution of both isomeric and non-isomeric human miRNA. The isomeric miRNA are structural isomers meaning they have the same nucleotide composition but the order in the sequence is different. There are 2,588 human microRNA present in this count (figure 3).

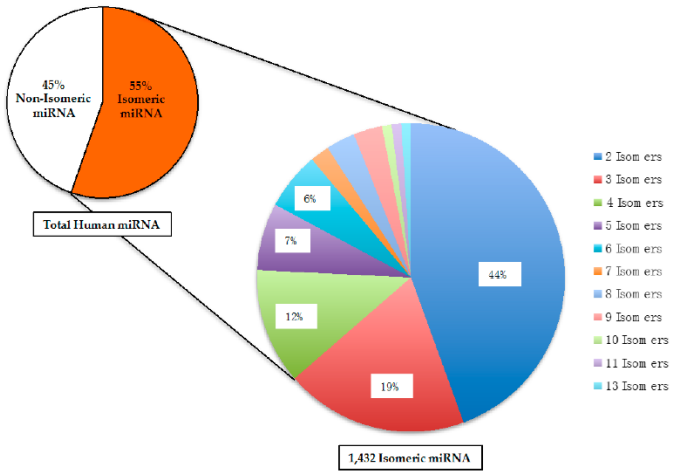

Figure 3. Total Human miRNA: Both Isomeric and Non-isomeric

The majority of the miRNA were in the group of isomers which had two isomers or (2 mers). Because of this, this project’s primary focus is on the SimiR

with two isomers as it provides a manageable size of isomers which depicts most of the data set.

There are many studies which associate human miRNA to different diseases. There have been reports of multiple miRNA associated to the same disease. Due to this, there exists a need for a quantitative and high thru-put method to detect miRNA.

1.1.6 RNA Modification and Detection Methods

Over 100 types of chemical modifications have been identified in cellular RNAs. Some of the important modifications to note are adenosine methylations, isomerization of uridine, ribose modifications, and cytosine modifications. The 5' cap modification in mRNA plays huge roles in metabolism and regulation. The most abundant internal mRNA modification is the N6-methlyladenosine in mammalian and long coding RNA[8]. There are initial beliefs that these modifications in these RNAs are used to stop the degradation and improve the gene performance within the cell. They are important for the stabilization of 2D and 3D structures. The purpose of these modifications can vary, and, when left unchecked, these modifications can be detrimental to both health and development. Different RNA modifications change the chemical composition of ribonucleic acid and have the potential to alter the function or stability of RNA[9]. The modifications are important as they affect the transcription of RNA by

altering base pairing, charges, secondary structure and/or the protein-RNA interactions[10,1].

The most studied and abundant modifications are in noncoding RNA. More specifically mRNAs, along with tRNAs, rRNAs, and snRNAs. Currently, there are a few proposed functions regarding the role of the modifications. There has been records of it impacting protein translations and localization, mRNA stability and different types of expression. In mammalian cells, there are mRNA present which have these methyladenine that helps sculpts the transcriptome. Through these processes, they help make up metabolism and represent new mechanisms which recognize gene expression[11].

Traditionally, there are several detection methods used to identify and validate different RNA modifications but there are limitations on each methodology. Some of the most common are: the bisulfite RNA sequencing (used for detection and quantification of m^5C); SCARLET (validation and quantification of m^6A and psi); and RiboMethSeq (Detection of quantification of 2'-O-Me)[12].

While these are well known and used methods, the limitations of each are important to note. While using the bisulfate RNA sequencing, there are many different errors that can occur during the experimentation. The Bisulfite sequencing uses deamination of the nucleotides to detect the m^5C modifications that are present in the sequencing. There is a high level of noise in the regions of

incomplete deamination of RNA. Furthermore, this method requires large quantities of sample. In this method, one can obtain single nucleotide resolution and can quantify the outcome, but there is only one modification which can be detected[12]. SCARLET, however is used to show modifications such as m^6A and psi and allow for information on the sequences. In SCARLET, there can only be a single site query and it is highly labor intensive and again, only shows for certain modifications. When comparing techniques such as PCR, one looks at all modifications present. Lastly, for the RibMethSeq method, the structure of the RNA can cause false positives. Also, abundant RNA species and high sequencing depth are needed to analyze the data. Because of the numerous limitations present, there is a need for a method that uses small amounts of material, is quick and cost effective.

Although a plethora of techniques exist which can be used to show the sequencing portion of the data, however, since the modifications are deleted, this is still a poor representation of the actual RNA samples. Few methods exist which can successfully detect the multiple RNA modifications present. In addition to this, there are minimal amounts of methods that can detect all the methods along with taking in account of the structure that the RNA possesses. Because we consider RNA structure to be synonymous to its function, it is important these are taken into account while using different detection methods.

Ion mobility mass spectrometry is the methodology that we propose to detect the modifications. Because there will be a quantification of the mass of the modifications. In addition to that, the ion mobility portion will allow for structural elucidation in the detection method.

1.1.7 Ion-Mobility Mass Spectrometry

We are proposing the use of ion mobility mass spectrometry to detect their presences of modifications and define structural differences in the isomeric RNAs. The mass spectrometry will be used to verify the presence of the RNA. The mass will show not only that the sequence has the right composition but also if there is anything in addition to the weight to indicate that a modification is present. The ion mobility will be used to show the different drift times present to help distinguish the difference between the isomers of the miRNA.

Ion mobility mass spectrometry (IM-MS) is a technique that has been used to determine the structures of small organic and inorganic molecules. The use of IM-MS has been used as a powerful structural elucidation tool[13]. This is an analytical technique that is used to separate an analyte by its mass and identify the ionized molecules based on their mobility in the gas buffer. In recent years, IM-MS has been used to determine structures of supra molecular assemblies that could not be accomplished by using X-ray crystallography or NMR[14]. The most novel part of this technique is understanding the drift time and how the shape of an ion affects its migration. First, the ion mobility portion will ionize the

different analytes. Then it will separate the ions according to their mobility through a buffer gas. Then, the mass spectrometer separates ions by their mass-to-charge ratio.

In recent studies ion mobility is becoming more heavily used for the analysis of biological materials in the fields of proteomics and metabolomics. Practical uses of IMS-MS include the detection of diseases such as lung cancer and pulmonary diseases[15]. It provides a higher resolution of the separations of protein fragments for analysis. In addition to the use of IMS, the collisional cross section values will be calculated to determine the presented drift time. Currently these CCS values of glycans and their fragments to help increase structural identification[16].

1.1.8 Ion-Mobility Drift Time

A drift time ion mobility spectrometer (DTIMS), is equipment which allows ions to move through an electric field in a drift tube in the presence of a noble gas molecule. There is an ionization region of the mass spectrometer that when heated causes the molecules to ionize and then sends them into a counter current flow of the neutral gas (nitrogen or helium). The non-ionized constituents in the sample are removed during this drift period. Through manipulation of the drift period, different ions can be separated in the mass spectrometer based on the size of the ion. Through the separation of the analytes by size, we propose

that we can separate the structural isomers through ion mobility and then validate by the mass spectrum.

1.1.9 Collisional Cross Sections

Collisional cross sections (CCS) values are those that are created while taking in account of how an object collides with gas molecules. It is defined as the area of an object that collides with another object through an interaction. This is used instead of a two-dimensional length because the analyte (the microRNA) will change shape as the ion mobility changes minorly while it is detected. Once external is energy introduced, the ion will rotate and create a three-dimensional shape. This number is important because it provides a prediction for how the analyte will respond in IM-MS. If the CCS values are drastically different, meaning that the shape is drastically different, then the drift time will also be different when placed into IM-MS.

If the object is large enough, the gas interaction would not need to be taken into account, but for small objects it is important. [17] Another noteworthy aspect of CCS values is how the object is struck. If an object is a perfect spear, then the way collides with another object is radially symmetric, it is simply a matter of hit or miss. However, once an object's concavity is considered, it is the difference between a solid collision and a swipe. It should be noted that molecules with the same size and makeup, but with differing concavities will have drastically different CCS values[18,19].

This data will later be used to create collision cross sections (CCS) of the structures which will later be used to allow us to look for different drift times in ion mobility mass spectrometry (IM-MS). This would provide a detection method that not only ensures that the structure is what it is believed to be but also to help differentiate the isomers. The collisional cross section is the area around the around the particle from the center that another particle must take in order for collision to occur. This is important because the CCS can be used to determine time of flight of the molecules based off of how the shape of the molecule will differ between different isomers.

1.2 Central Hypothesis and Objectives

To address these issues for RNA research work as well as improving the accuracy on identifying specific RNA biomarkers, this study **aims** to investigate some of the intrinsic properties of RNA biomarkers that may affect the ion mobility mass spectrometric measurements of RNA samples. We **hypothesize** the differences in RNA folding among isomeric RNA molecules are resolvable by using the current technologies in ion mobility mass spectrometry.

The **long-term goal** of this study is to develop an analytical method for differentiating and identifying isomeric RNA with or without any modification base on the RNA folding in the corresponding molecular ions in the gas phase. Particularly, we are interested in the identification of specific isomeric RNA biomarkers which have identical nucleotide composition and high sequence

similarity. MicroRNA (miRNA) are selected as an initial model in this study, and possible folding are simulated by using various computer modeling programs.

CHAPTER II

EXPERIMENTAL

2.1 Methods

2.1.1 The mFold Web Server Procedure

The software is from the mfold Web Server from The RNA Institute College of Arts and Sciences at the University at Albany State University of New York. For all portions of this experiment, RNA Folding Form (version 2.3 energies) will be used to determine the two-dimensional folding. This webserver was accessed on January 23, 2017.

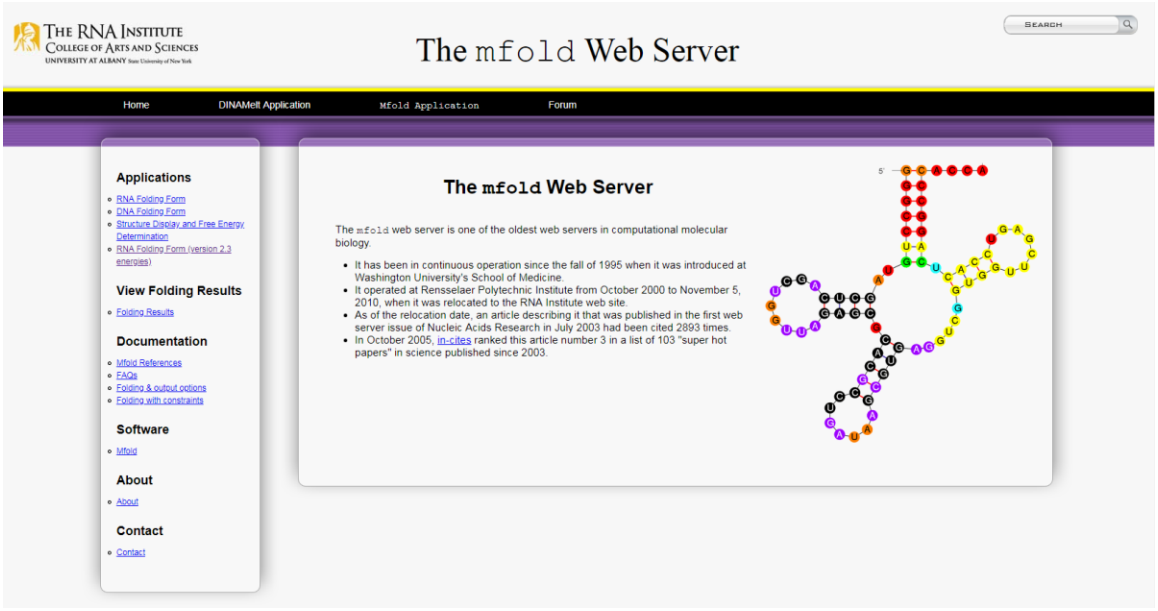

Figure 4. The Homepage for The mfold Webserver

Before beginning the experiment, the software needed to be validated with known tRNA structures. These were structures that had previously been validated through x-ray-crystallography. It is important to note that unlike the experimental structures that were tested at 25℃, this particular experiment was simulated at 37ºC to keep with the biological conditions that the tRNA are found (fig. 6).

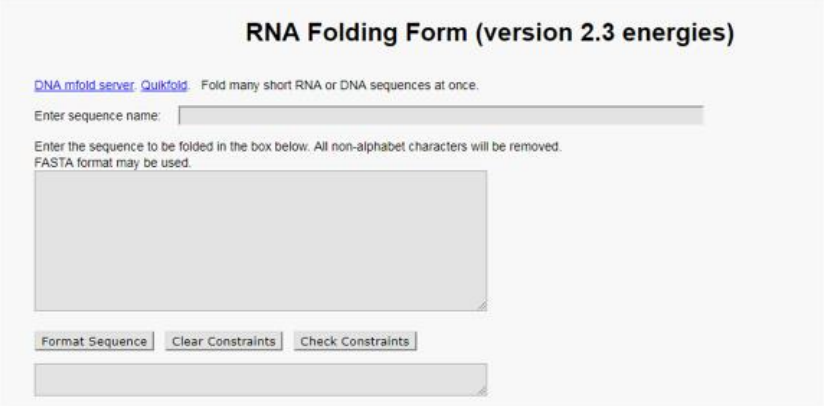

Figure 5. RNA Folding Sequence Loading Dock

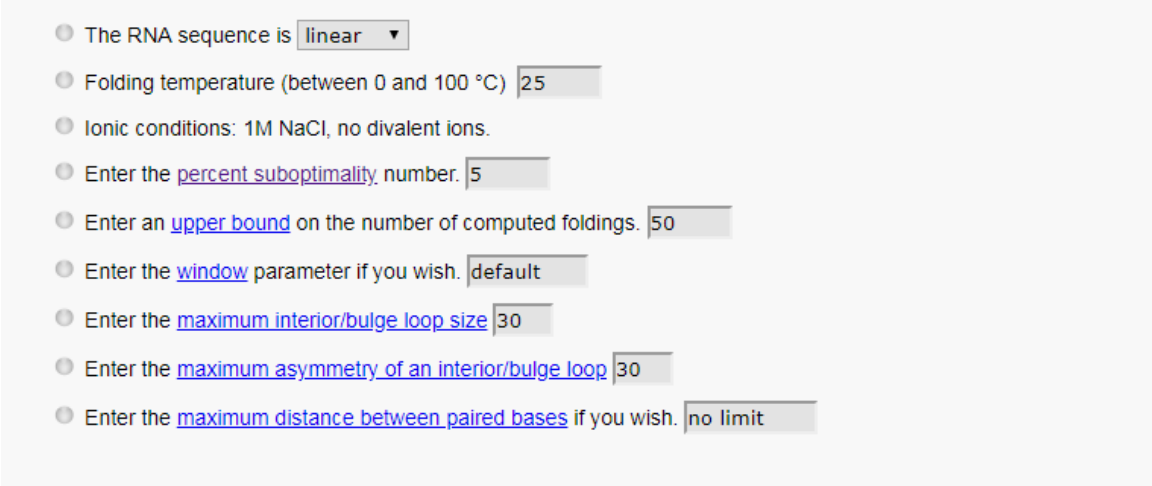

Figure 6. The mfold Webserver Parameters

The tRNA structures that were analyzed were 1YFG and 2TRA. These two structures not only had x-ray structures but possessed short sequences. The sequences were added and then they were formatted. The rest of the settings were left as a default to maintain biological relevance. In the case of 2TRA, three possible structures were outputted but all looked relatively similar to one another. In addition to that, their energies were all similar to one another and all resembled the x-ray structure. The primary difference was the wrapping of the tails at the end. This major drawback is because this software only produces two dimensional figures, the size of the actual loop and the angles that this hairpin loops are easily distorted. Also, this software will allow for non-Watson Crick base pairs. *The mfold Web Server* will just rank the non- Watson Crick base pair as a very weak bond.

Using *The mfold Web Server*, the folding of 632 structural isomers of microRNA (SimiR) were simulated and their thermodynamic parameters were determined. The RNA sequence was copied into *mfold*. “RNA Folding Form (version 2.3)” and the following parameter was adjusted: temperature was set to 25 ºC.

This specific one was chosen because it not only gave us more flexibility of changing the parameters of the microRNA to reflect that in a room temperature at which the future measurements will be carried out. Once the parameters were set, projected folding and thermodynamic information (ΔG, ΔH, ΔS, and T_m were recorded and used for later analysis.

2.1.2 SimRNAweb Server Procedure

First we needed to validate the software.The software used to determine some of the three-dimensional folding for the miRNA is SimRNAweb. This software was accessed on May 20, 2017.

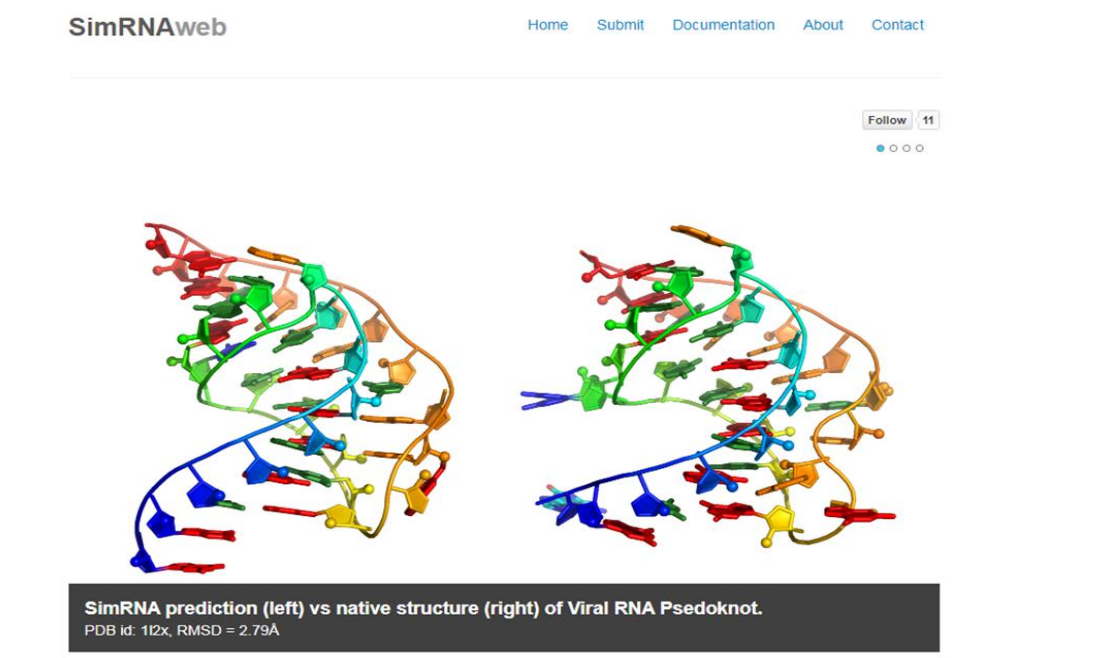

Figure 7. The SimRNA Homepage

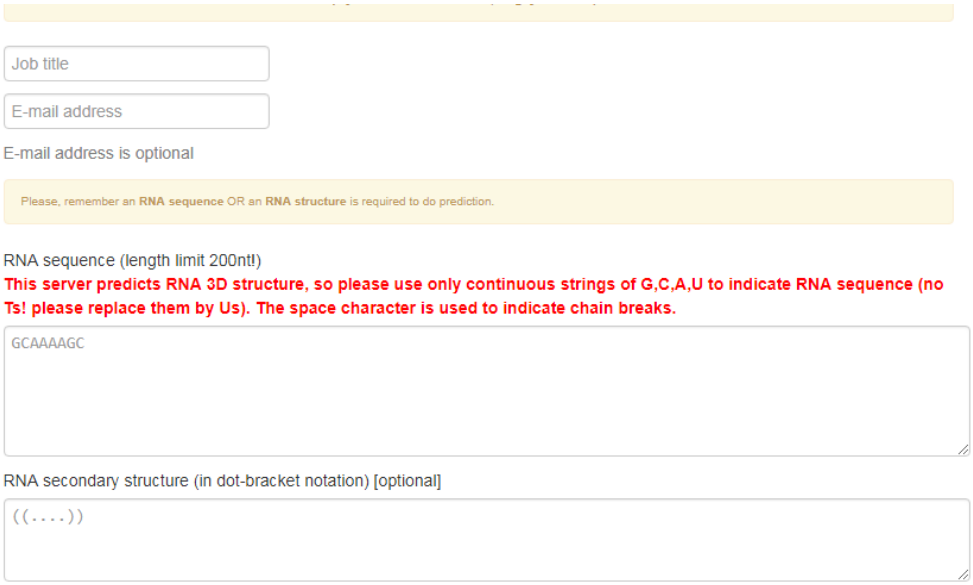

Figure 8. SimRNA Job Loading Station

Before beginning the experiment, the software needed to be validated with known tRNA structures. These were structures that had previously been validated with a crystal structure. The sequences were taken into *SimRNA* and preformed through an unguided search. That means no PDB file was added and there was no dot-bracket notation to lead the software into a tertiary structure. *SimRNA* claimed to use preexisting known structures so the structure should have come out similar to the actual one. The tRNA structures that were analyzed. The software output 5 structures for every sequence added. The two tRNA structures that were used were 1YFG and 2TRA. Both tRNA had an output of the strongest five possible structures. When observing the most stable five structures of 1YFG, the fifth was almost identical to the one predicted through x-

ray crystallography. In the case of 2TRA, the first was most identical to that predicted through x-ray crystallography.

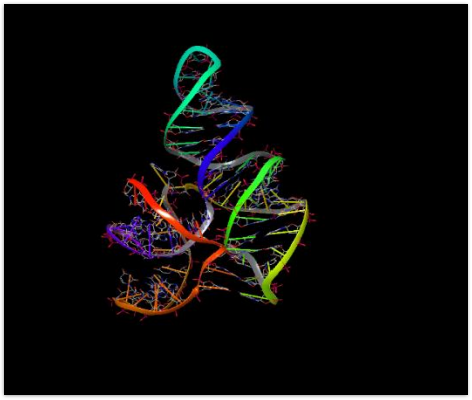

Figure 9. The Validated 3D tRNA Structure 2TRA

There were two important outcomes to these experiments. One is that that this software produced results on par with the current advances on computational structures. The second is that although they have a ranking system, the first structure does not guarantee that is the best structures, thus all structures produced must be analyzed.

2.1.3 Vfold3D Procedure

We first needed to validate this software. Vfold3D was also used in this experiment. The structures were given a name and presented in dot-bracket notation. This software did not allow for any additional parameters to be set as it was based off the premise of having sound knowledge of the structure prior to folding. Because of this, the validation of this software strictly relies on the validation of the *mfold* software and precision of the dot-bracket notation. There

is no free reign for software to insure fast folding. The following strains were simulated: hsa-miR-892c-5p, hsa-miR-3117-3p, hsa-miR-95-5p, hsa-miR-8080, hsa-miR-4726-5p, hsa-miR-671-5p, hsa-miR-503-5p.

2.1.4 Sybyl Procedure

All of the structures that were collected from *SimRNA* and *Vfold3D* will be taken over into *Sybyl*. From here there will be further minimizations preformed. A Kollman Force Field with be applied with AMBER as the charge. This will be selected because Kollman Forces Fields are used for nucleic acids and will apply the restrictions that nucleic acids possess. The AMBER will place a charge throughout its backbone to show real life settings. Once all minimizations have been completed, the structures will be placed into x, y, z format and collisional cross sections will be calculated.

2.1.5 IMoS Procedure (CCS Values)

For this portion IMoS will be used to calculated CCS values. The software was downloaded on March 22, 2018. The software was downloaded in addition to a MATLAB capability package.

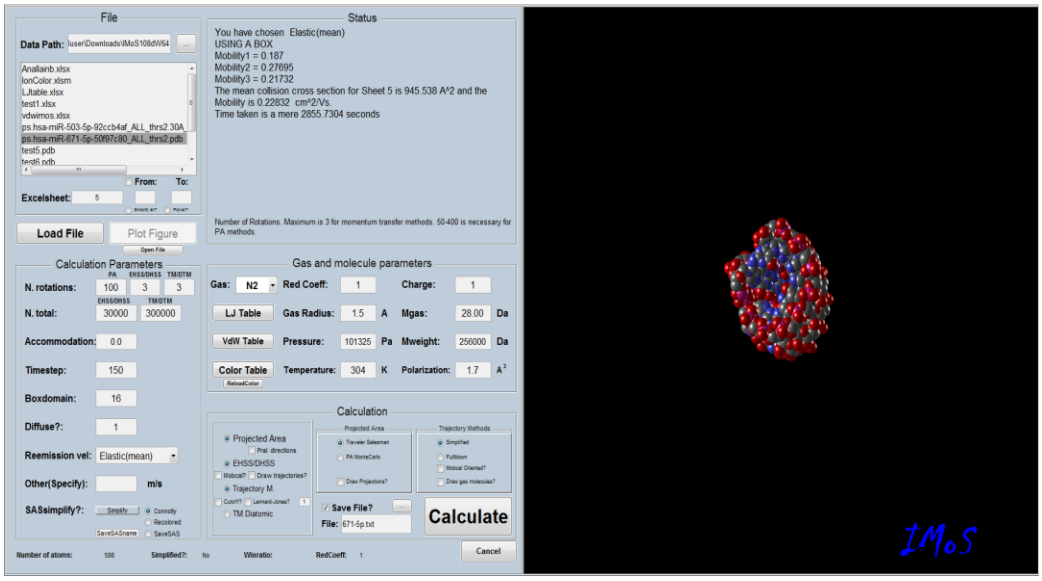

Figure 10. Loading and Plotting Screen from IMoS

From there, the pre-minimized structures are saved into the folders that are added after the download. The structures have to have the charges present for this software to acknowledge and run the CCS values. The structure will be loaded, plotted. This is to insure the structures are the same and no charges have been made while loading. Then calculated at the ideal parameters. The runs will take several hours to complete due to the size of the structures.

CHAPTER III

RESULTS AND CONCLUSION

3.1 Results

3.1.1 Mfold Results

Table 1. Representing Data from *mfold* Results of 623 Isomers

Results	MiRNA	Sequence (5' to 3')	ΔG	ΔH	ΔS	T_m
Unfolded	hsa-miR-1238-3p	CUUCCUCGUCUGUCUGCCCC	0.08 kcal/mol	-18.60 kcal/mol	-62.65 cal/Kmol	23.7°C
Unfolded	hsa-miR-4428	CAAGGAGACGGGAACAUGGAGC	-0.34	-14.00	-45.82 cal/Kmol	32.4°C
Folded	hsa-miR-4726-5p	AGGGCCAGAGGAGCCUGGAGUGG	-7.31 kcal/mol	-51.50 kcal/mol	-148.21 cal/Kmol	74.3°C
Folded	hsa-miR-671-5p	AGGAAGCCCUGGAGGGGCUGGAG	-11.78 kcal/mol	-64.70 kcal/mol	-177.49 cal/Kmol	91.4°C
When calculating the values, the errors were ± 5%, ± 10%, ± 11% and 2-4C for free energy, enthalpy, entropy and Tm, respectively.						

The isomers underwent an initial filtering process only looking at the differences in their ΔG, as seen in table 1. In addition, there is a representing two-dimensional projected folding of the images (fig. 10).

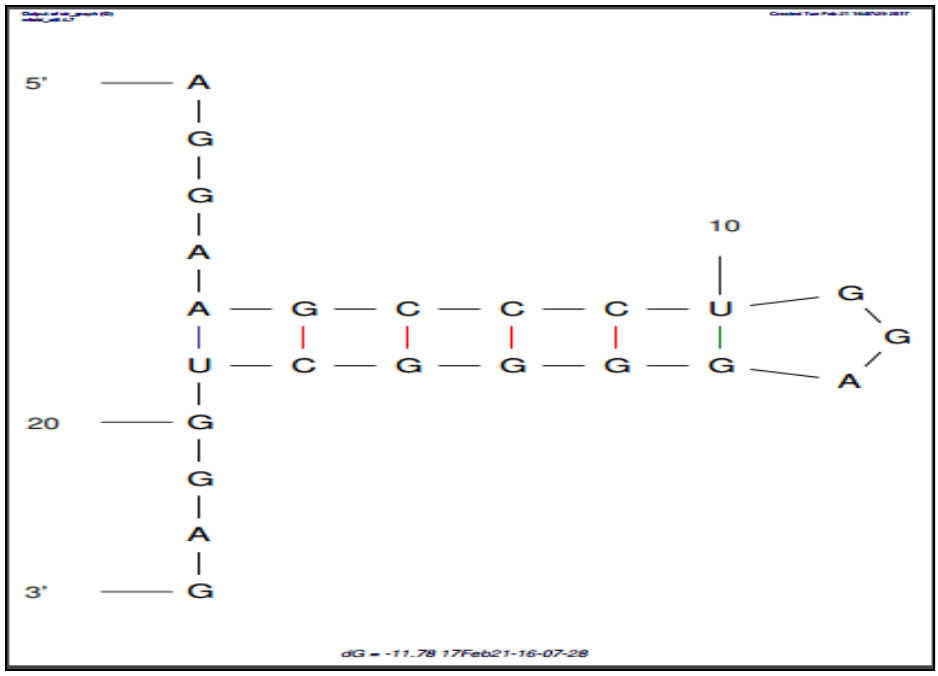

Figure 11. A 2D Structure from SimiR hsa-miRNA-671-5p

The initial screening showed many different folding results for most of the miRNA. For one sequence, *mfold* would provide anywhere between one to eight different folding results for each individual miRNA. It was important to see the frequency of the folding. Initially, it was important to negate the fact that they were isomers just to see how many times that *mfold* would give us a structure with only one possible folding shown in figure 11.

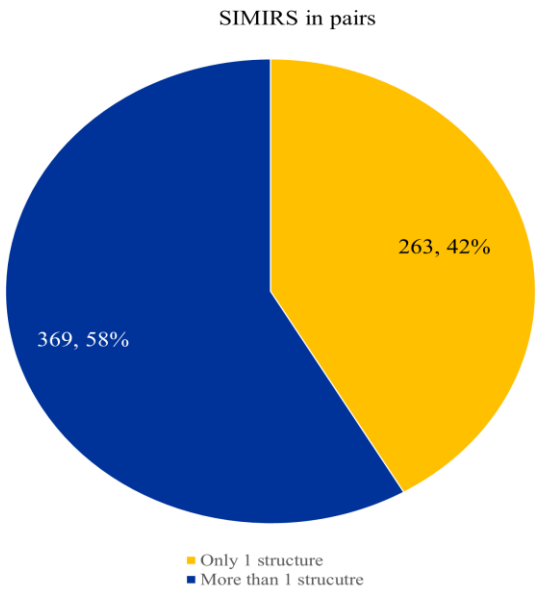

Figure 12. Sequences that Have Only One Possible Folding

Once there was strong evidence that many sequences, despite being isomeric, it was seen that this happened about 42% of the total population of the isomeric RNA as shown in figure 11. The isomeric pairs that both only had one structure were separated. From the 342 pairs, it was shown that 17% showed only one structure through the parameters set in mfold.

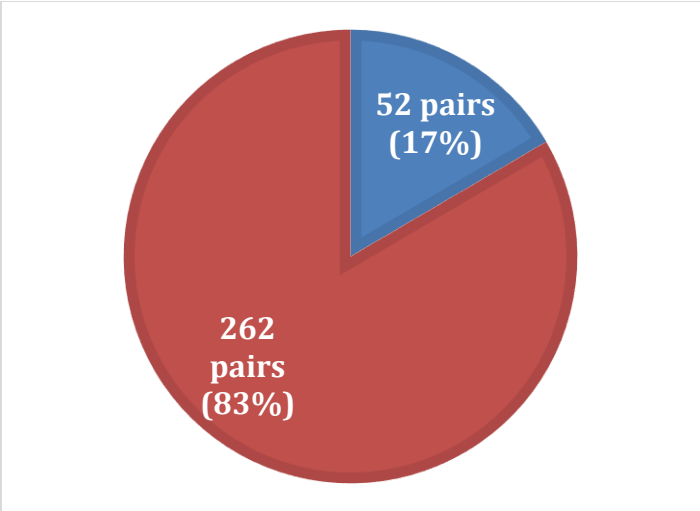

Figure 13. SimiR with Only One Structure

In addition to only seeing one result following folding, we looked at the structures that were considered to have one that one folded and one isomer that is unfolded. This would have to have a ΔG value that is about a 2 kcal/mol difference.

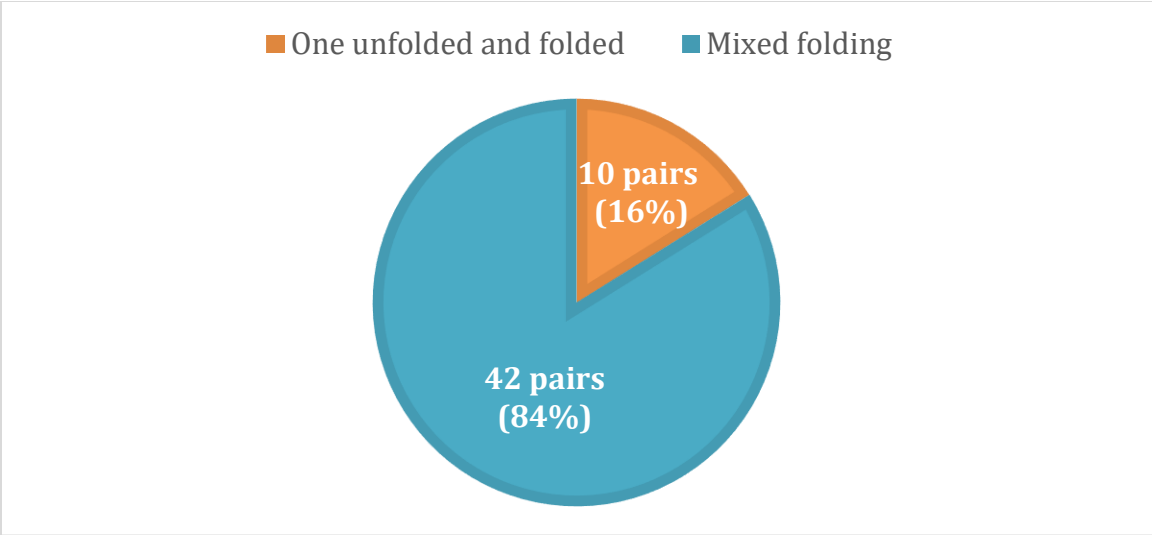

Figure 14. Isomeric Pairs: One Folded and Unfolded

The table shows the structures that were demined to have one folded and one that unfolded isomer.

Table 2. A Sample of Selected Isomers and Varying ΔG

	(kcal/mol)	Differences in ΔG			(kcal/mol)	Differences in ΔG
hsa-miR-4726-5p	-7.31	4.47		**hsa-miR-4428**	-0.34	3.06
hsa-miR-671-5p	-11.78			**hsa-miR-4667-5p**	-3.4	
hsa-miR-1264	-5.64	2.40		**hsa-miR-95-5p**	-5.66	3.17
hsa-miR-4666b	-3.24			**hsa-miR-369-3p**	-2.43	
hsa-miR-503-5p	-9.5	4.00		**hsa-miR-892c-5p**	-1.74	3.39
hsa-miR-8080	-5.5			**hsa-miR-3117-3p**	-5.13	
hsa-miR-1197	-7.22	2.47		**hsa-miR-6812-3p**	-1.11	2.98
hsa-miR-182-3p	-4.75			**hsa-miR-6800-3p**	-4.09	

The Δ G values and the ΔH values allowed for the stability of the predicted folding to be predicted. The values that were projected for the Δ G which was the primary thermodynamic information was that used to separate and narrow down the isomers. The project started with over 600 microRNA, so there needed a way to determine which structures should be further tested. All Δ G values were examined and cut offs were created to analyze the data. The ranges that were obtained in the Δ G values were between 0.22 kcal/mol to -11.78 kcal/mol. Those with lower than -4 kcal/mol were the cut off on a stable folding and anything that was higher than that were determined linear, although those that were around -3 kcal/mol were seen as undifferentiable.

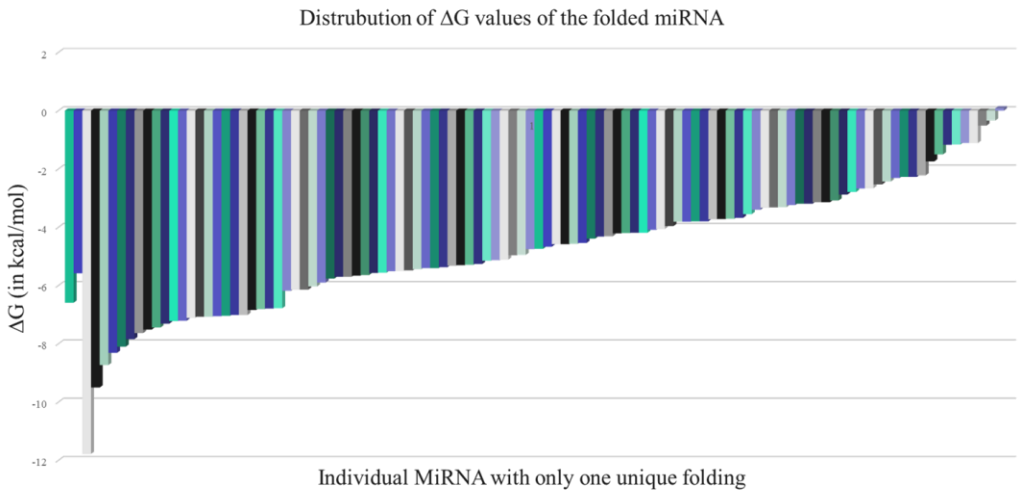

Figure 15. Varying ΔG Values for the miRNA that Only Have One Folding

Because these were isomers, it was important to pick pairs that had a large enough difference in Δ G values from one another to ensure that the

structures were as differently as initially believed and had different stabilities. The T_m values were also recorded. Because the Collisional Cross Sections was the focus of this project, the melting point of each miRNA will be used later for further mass spectrometry work. It will allow us to know how to adjust the voltage added to insure different drift times amongst the isomers.

Limitations: One of the most significant limitations of this software was the ionic charge that was included in the structure. The software would not allow for folding with this charge even though the temperature of the environment was able to be changed. The important aspect to note however is that the same error was introduced to all the microRNA.

The importance of this step was to see if the SimiR could fold. On top of checking for different folding in these RNA, it was important to note if the isomers folded differently from one another. Because these RNA were so small, and differed by only one or two base pairs, there was large uncertainties in the SimiR folding.

This confirmed that computational models can predict stable folding of RNA. Further software was used to show a three-dimensional folding for the RNA with the strongest folding, and with the lowest Δ G values.

3.1.2 SimRNAweb Results

The strains were simulated: hsa-miR-892c-5p, hsa-miR-3117-3p, hsa-miR-95-5p, hsa-miR-8080, hsa-miR-4726-5p, hsa-miR-671-5p, hsa-miR-503-5p. These simulations were performed both as guided and unguided simulations. In the absence of guidance, there were some differences that were noted in all the structures. In the guided simulations, the RNA sequence length was provided and then the additional restrains were added from the *mfold* software. The secondary structure was placed in dot-bracket notation and they were submitted for the job. The jobs were also submitted without any restrains for a comparison. There were 500 steps in each simulation.

Figure 16. Isomer hsa-miR-671-5p from mfold to SimRNA Web

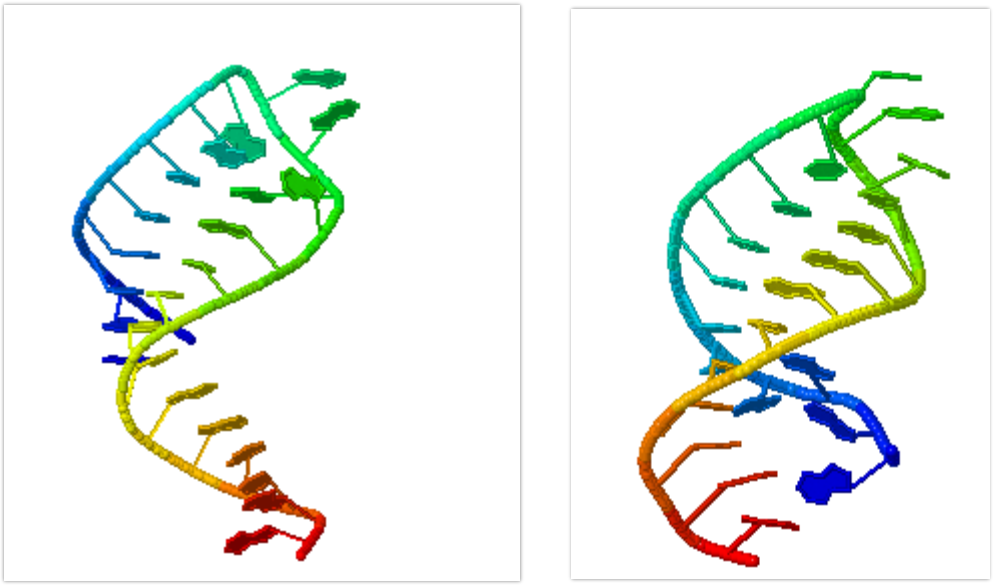

Figure 17. Sample 3D Structures from simRNA: hsa-miR-4726-5p and hsa-miR-671-5p

3.1.3 Vfold3D Results

Once these structures were generated, a quick comparison showed the similarities between the structures from the two-different software. Even when simulations were ran on different engines, they were identical in shape. Because *SimRNA* has more freedom in folding, it was important to note that there were some differences between them and the *mfold* structures. In fact, in some instances, the top three preferred structures were drastically different from the two-dimensional folded option. Regards of the differences in the isomers, they were kept for future analysis. It is also important to note that this software does not allow non-Watson Crick base pairing

*For each software, it is important to note that they had to be placed into dot bracket notation.

3.1.4. Sybyl Results

Once the structures were uploaded and minimized in Sybyl, the outputs were screened and minimized, the structures were placed into x,y,z, formatted to place into IMoS to be calculated. All partial charges are included in the structures to allow it to run for the CCS values. A sample of the minimized structures is found in Figure 17.

Figure 18. hsa-miR-4726-5p and hsa-miR-671-5p minimized on Sybyl

3.1.5. IMoS Results

From this software we are able to calculate the CCS values for each of the SimiR. In addition to that, the mobilities were calculated.

Table 3. Sample Data from IMoS

	CCS values (Å²)	Linear CCS values (Å²)
hsa-miR-4726-5p	1044	1662
hsa-miR-671-5p	945	1690

Table 3 shows some of the sample calculations for the CCS values that were calculated in IMoS. This shows that in their theoretically folded states, the CCS values were rather different from one another. Isomer hsa-miR-4726-5p has a value of 1004A while the msa-miR-671-5p has on of 945. This gives a resolution of 10 which for macromolecules that are isomers, are a good natural start.. The other column gives a simulation of what it could be if the isomer was melted and made to lay linear. If this is the case, values are much higher than if not and then the resolution and separation could be better displaced in ion mobility. Although these are theoretical, they show promise into increasing the difference are in fact resolvable in ion mobility mass spectrometry.

3.2 Conclusion

Through this project we showed that through the use of theoretical computational methods, we can predict the probable folding of miRNA in the gas phase. In addition to this we were able to take these folding from a two-dimensional space and form three-dimensional folding. From there we calculated

predicated CCS values of each of this SimiR. With that, we are able to theoretically conclude that these values are different enough to resolve in Ion Mobility Mass Spectrometry.

REFERENCES

(1) Feature, T. (2004) THE RNA CODE COMES B Y K E L LY R A E C H I.

(2) Sedova, A., and Banavali, N. K. (2016) RNA approaches the B-form in stacked single strand dinucleotide contexts. *Biopolymers 105*, 65–82.

(3) Hermann, T., and Patel, D. J. (2000) RNA bulges as architectural and recognition motifs Thomas Hermann and Dinshaw J Patel RNA bulges constitute versatile structural motifs in the. *Structure 8*, 47–54.

(4) Mwangi, J., and Chiu, N. (2016) High Percentage of Isomeric Human MicroRNA and Their Analytical Challenges. *Non-Coding RNA 2*, 13.

(5) Wirta, V. (2006) Mining the transcriptome methods and applications Valtteri Wirta. *Stem Cells*.

(6) Cooper, G. M., and Hausman, R. E. (2004) The cell: A molecular approach 713.

(7) Song, J., and Yi, C. (2017) Chemical Modifications to RNA: A New Layer of Gene Expression Regulation. *ACS Chem. Biol. 12*, 316–325.

(8) Liu, N., Parisien, M., Dai, Q., Zheng, G., He, C., and Pan, T. (2013) Probing N 6 -methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA Probing N 6 -methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. *Rna 19*, 1848–1856.

(9) Lu, J., Getz, G., Miska, E. A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B. L., Mak, R. H., Ferrando, A. A., Downing, J. R., Jacks, T., Horvitz, H. R., and Golub, T. R. (2005) MicroRNA expression profiles classify human cancers. *Nature 435*, 834–838.

(10) Roundtree, I. A., Evans, M. E., Pan, T., and He, C. (2017) Dynamic RNA Modifications in Gene Expression Regulation. *Cell 169*, 1187–1200.

(11) Yue, Y., Liu, J., Cui, X., Cao, J., Luo, G., Zhang, Z., Cheng, T., Gao, M., Shu, X., Ma, H., Wang, F., Wang, X., Shen, B., Wang, Y., Feng, X., He, C., and Liu, J. (2018) VIRMA mediates preferential m6A mRNA methylation in 3′UTR and near stop codon and associates with alternative polyadenylation. *Cell Discov. 4*, 10.

(12) Helm, M., and Motorin, Y. (2017) Detecting RNA modifications in the epitranscriptome: Predict and validate. *Nat. Rev. Genet. 18*, 275–291.

(13) Bleiholder, C. (2015) A local collision probability approximation for predicting momentum transfer cross sections. *Analyst 140*, 6804–6813.

(14) Ruotolo, B. T., Hyung, S. J., Robinson, P. M., Giles, K., Bateman, R. H., and Robinson, C. V. (2007) Ion mobility-mass spectrometry reveals long-lived, unfolded intermediates in the dissociation of protein complexes. *Angew. Chemie - Int. Ed. 46*, 8001–8004.

(15) Pereira, J., Porto-Figueira, P., Cavaco, C., Taunk, K., Rapole, S., Dhakne, R., Nagarajaram, H., and Câmara, J. S. (2015) Breath analysis as a potential and non-invasive frontier in disease diagnosis: An overview. *Metabolites 5*, 3–55.

(16) Aizpurua-Olaizola, O., Sastre Toraño, J., Falcon-Perez, J. M., Williams, C., Reichardt, N., and Boons, G. J. (2018) Mass spectrometry for glycan biomarker discovery. *TrAC - Trends Anal. Chem. 100*, 7–14.

(17) Wyttenbach, T., Bleiholder, C., and Bowers, M. T. (2013) Factors contributing to the collision cross section of polyatomic ions in the kilodalton to gigadalton range: Application to ion mobility measurements. *Anal. Chem. 85*, 2191–2199.

(18) Wyttenbach, T., Bleiholder, C., Anderson, S. E., and Bowers, M. T. (2015) A new algorithm to characterise the degree of concaveness of a molecular surface relevant in ion mobility spectrometry. *Mol. Phys. 113*, 2344–2349.

(19) Goldstein, M., Zmiri, L., Segev, E., Wyttenbach, T., and Gerber, R. B. (2014) An atomistic structure of ubiquitin +13 relevant in mass spectrometry: Theoretical prediction and comparison with experimental cross sections. *Int. J. Mass Spectrom. 367*, 10–15.